



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/752,432	01/07/2004	Jason G. Maxham	53217-014	3709

7590

06/13/2005

MCDERMOTT, WILL & EMERY
600 13th Street, N.W.
Washington, DC 20005-3096

EXAMINER

LIANG, GWEN

ART UNIT PAPER NUMBER

2162

DATE MAILED: 06/13/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/752,432	MAXHAM ET AL.	
	Examiner	Art Unit	
	GWEN LIANG	2162	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 21 April 2005.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-34 is/are pending in the application.
- 4a) Of the above claim(s) 14-28 is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-13 and 29-34 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 07 January 2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|---|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |



8.000

DETAILED ACTION

1. This action is responsive to communications through the applicant's amendment, filed on 04/21/2005.

Claim Rejections - 35 USC § 112

2. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

3. Claim 34 is rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

In independent claim 34, the claimed subject matter "said computer using an operating system that stores electronic documents substantially equally throughout the cluster" is not described in the specification in the same way as in the claim. For example in the specification page 4, [14], lines 2-3, "said computer using an operating system that stores electronic documents in a hard disk drive throughout the cluster"; clearly illustrates the difference between the description in the specification and what is claimed in the applicant's invention.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 1, 4-13, and 29-33 are rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900), and further in view of Getchius et al., "Getchius " (U.S. Patent No. 6,493,721).

With respect to claim 1, Burrows discloses a method ... comprising:

a) determining a file type for each native document of the plurality of native documents (col. 7, lines 58-65, "For example, the page 200 of FIG. 4 can have associated page attributes 250. Page attributes 250 can include □ADDRESS□ 251, □DESCRIPTION□ 252, □SIZE□ 253, □DATE□ 254, □FINGERPRINT□ 255, □TYPE□ 256, and □END_PAGE□ 257, for example. The symbol "□," represents one or more characters which cannot be confused with the characters normally found in words, for example "space," "underscore," and "space" (sp_sp)"; col. 8, lines 24-25, "The TYPE attribute 256 may distinguish pages having different multimedia content or formatting characteristics"; Figure 4, element 256);

b) creating a fingerprint for each native document (col. 8, lines 16-23, "The FINGERPRINT 255 represents the entire content of the page. The fingerprint 255 can be produced by applying one-way polynomial functions to the digitized content.

Typically, the fingerprint is expressed as an integer value. Fingerprinting techniques ensure that duplicate pages having identical content have identical fingerprints. With very high probabilities, pages containing different content will have different fingerprints.”);

c) de-duplicating each native document in accordance with the fingerprint (col. 1, lines 42-45, “Therefore, it is desired to provide a technique which minimizes the likelihood that duplicate pages are indexed. The technique should also allow for reindexing as duplicate pages are deleted.”; col. 2, lines 41-42, “FIG. 24 shows a process for detecting duplicate pages; FIG. 25 is a flow diagram of a process for deleting pages;”; col. 5, lines 12-14, “The maintenance module 80 also effectively deals with duplicate Web pages containing substantially identical content.”);

d) extracting data from each native document; e) associating extracted data with a corresponding native document (col. 5, lines 33-38, “A page 200 can be defined as a data record including a collection of portions of information or “words” having a common database address, e.g., a URL. This means that a page can effectively be a data record of any size, from a single word, to many words, e.g., a large document, a data file, a book, a program, or a sequence of images.”; col. 11, line 66 – col. 12, line 7, “The samples are used to generate summary entries 925 in the second level summary data structure 72. Each summary entry 925 includes the word 926 associated with the sample, and the sampled location associated with the word. In addition, the summary entry 925 includes a pointer 928 of the next entry in the compressed data structure 71 following the sampled entry. The summary data structure 72 can also be mapped into

Art Unit: 2162

fixed size blocks or disk files to fully populate the summary data structure 72.", wherein the words are extracted from the native documents); and

f) distributing the plurality of native documents and extracted data amongst a plurality of nodes of the document management computer system (col. 1, lines 65-67, "FIG. 1 is a block diagram of a distributed database storing multimedia information indexed and searched according to the invention;").

However Burrows does not explicitly disclose that the distribution of the plurality of native documents and extracted data amongst a plurality of nodes is substantially equal.

Getchius teaches a method comprising distributing data substantially equally amongst a plurality of nodes (Abstract, "The system for performing online data queries is a distributed computer system with a plurality of server nodes each filly redundant and capable of processing a user query request.").

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a method of distributing data substantially equally amongst a plurality of nodes as disclosed by Getchius into the method of distributing the plurality of native documents and extracted data amongst a plurality of nodes of the document management computer system as disclosed in Burrows so that each node is capable of responding to any search request (col. 18, lines 41-43). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

Claim 4 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a method wherein step (c) further comprises comparing the fingerprint of each native document with a plurality of fingerprints comprised of the fingerprints for each native document to be uploaded (col. 28, lines 40-47).

Claim 5 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a method wherein step (c) further comprises comparing the fingerprint of each native document with at least one fingerprint corresponding to a native document stored in the document management computer system (col. 28, lines 40-47).

Claim 6 is rejected for the reasons set forth hereinabove for claim 4 and furthermore Burrows discloses a method comprising discarding native documents that are determined to be the same in accordance with the comparison of fingerprints (Title; col. 1, lines 42-45; col. 8, lines 16-23).

Claim 7 is rejected for the reasons set forth hereinabove for claim 5 and furthermore Burrows discloses a method comprising discarding native documents that are determined to be the same in accordance with the comparison of fingerprints (Title; col. 1, lines 42-45; col. 8, lines 16-23).

Claim 8 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a method wherein step (d) further comprises creating at

least one data file corresponding to the extracted data for each native document (col. 11, line 66 – col. 12, line 7).

Claim 9 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a method wherein step (d) further comprises creating a plurality of data files corresponding to the extracted data for each native document (col. 11, line 66 – col. 12, line 7).

Claim 10 is rejected for the reasons set forth hereinabove for claim 9 and furthermore Burrows discloses a method wherein the plurality of data files includes files selected from a group consisting of a text file, a meta data file, an XML file and a HTML file (col. 8, line 66 – col. 9, line 8).

Claim 11 is rejected for the reasons set forth hereinabove for claim 10 and furthermore Burrows discloses a method wherein in step (e), a data table is created for at least one native document for defining an association with the plurality of data files (col. 14, lines 35-40).

Claim 12 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a method wherein in step (e), a data table is created for at least one native document for defining an association with extracted data (col. 14, lines 35-40).

Claim 13 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Getchius discloses a program product, comprising executable code

Art Unit: 2162

transportable by at least one machine readable medium, wherein execution of the code by at least one programmable computer causes the at least one programmable computer to perform a sequence of steps, comprising the steps recited in claim 1 (col. 19, line 52 – col. 20, line 8).

Claim 29 is rejected for the reasons set forth hereinabove for claim 1 and furthermore Burrows discloses a system comprising a computer in communication with the plurality of computer nodes for receiving a plurality of input files to be uploaded to the plurality of computer nodes (col. 2, lines 51-56, "FIG. 1 shows a distributed computer system 100 including a database to be indexed. The distributed system 100 includes client computers 110 connected to server computers (sites) 120 via a network 130. The network 130 can use Internet communications protocols (IP) to allow the clients 110 to communicate with the servers 120.").

The subject matter of claims 30 and 33 are rejected in the analysis above in claim 1, and therefore these claims are rejected on that basis.

The subject matter of claims 31 and 32 are rejected in the analysis above in claims 8 and 10 respectively, and therefore these claims are rejected on that basis.

6. Claim 2 is rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900), further in view of Getchius et al., "Getchius " (U.S. Patent No. 6,493,721), and further in view of Okabe et al., "Okabe " (U.S. Publication No. 2001/0025287)

Claim 2 is rejected for the reasons set forth hereinabove for claim 1. However the combination of Burrows and Getchius does not explicitly teach a method comprising the step of extracting native document(s) included in the plurality of documents from an archive file.

Okabe teaches the step of extracting native document(s) included in the plurality of documents from an archive file (page 6, section [0077]).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a step of extracting native document(s) included in the plurality of documents from an archive file as disclosed by Okabe into the method of managing a plurality of native documents as disclosed in the combination of Burrows and Getchius. The motivation obviously is to obtain documents from the archive through the extraction (page 6, section [0077]). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

7. Claim 3 is rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900), further in view of Getchius et al., "Getchius " (U.S. Patent No. 6,493,721), . and further in view of Zabetian (U.S. Publication No. 2001/0011350).

Claim 3 is rejected for the reasons set forth hereinabove for claim 1. However the combination of Burrows and Getchius does not explicitly teach a method wherein the fingerprint for each native document is created using a MD5 checksum.

Zabetian teaches a method wherein the fingerprint for each native document is created using a MD5 checksum (page 4, section [0037]).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a method wherein the fingerprint for each native document is created using a MD5 checksum as disclosed by Zabetian into the method of creating a fingerprint for each native document as disclosed in the combination of Burrows and Getchius, where a tamper proof checksum algorithm is desired, MD5 with DES encryption can be used (MD5-DES) (page 4, section [0037]). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

8. Claim 34 is rejected under 35 U.S.C. 103(a) as being unpatentable over Burrows (U.S. Patent No. 5,745,900), further in view of Narendran et al., "Narendran " (U.S. Patent No. 6,070,191), and further in view of Froessl (U.S. Patent No. 5,444,840).

With respect to claim 34, Burrows discloses a system ... comprising:
a PC type computer connected in a parallel cluster (col. 2, lines 51-56),
said computer using an operating system that stores electronic documents throughout the cluster (col. 3, lines 1-4 & 34-44; col. 11, lines 38-44; col. 15; lines 11-14, "This would be the case where the database indexed, the client programs, the search engine 140, and the index 70 all reside on a single computer system, e.g., a PC or workstation."; col. 1, lines 65-67, "FIG. 1 is a block diagram of a distributed database storing multimedia information indexed and searched according to the invention;"),
said operating system generating a fingerprint for each document (col. 8, lines 16-23, "The FINGERPRINT 255 represents the entire content of the page. The fingerprint 255 can be produced by applying one-way polynomial functions to the

digitized content. Typically, the fingerprint is expressed as an integer value.

Fingerprinting techniques ensure that duplicate pages having identical content have identical fingerprints. With very high probabilities, pages containing different content will have different fingerprints.”);

where each document is identified by its file extension (col. 7, lines 58-65, “For example, the page 200 of FIG. 4 can have associated page attributes 250. Page attributes 250 can include □ADDRESS□ 251, □DESCRIPTION□ 252, □SIZE□ 253, □DATE□ 254, □FINGERPRINT□ 255, □TYPE□ 256, and □END_PAGE□ 257, for example. The symbol “□,” represents one or more characters which cannot be confused with the characters normally found in words, for example “space,” “underscore,” and “space” (sp_sp); col. 8, lines 24-25, “The TYPE attribute 256 may distinguish pages having different multimedia content or formatting characteristics”; Figure 4, element 256; It is well known to an ordinary skill in the art that each file has a file extension, which indicate the type of the file or document);

and given a unique identification number (col. 26, lines 4-6, “Each entry 2201 includes an identification (page_id) 2210 of a qualified page”),

each of a plurality of documents having at least one of either meta-data, text or attachments that are indexed for web-based retrieval from the cluster (col. 8, line 66 – col. 9, line 8, “Attribute values or metawords can be generated for portions of a page. For example, the words of the field 230 may be the “title” of the page 200. In this case the “title” has a first word 231 and a last word 239. In “html” pages, the titles can be expressly noted. In other types of text, the title may be deduced from the relative

Art Unit: 2162

placement of the words on the page, for example, first line centered. For titles, the parsing module 30 can generate a "BEGIN_TITLE" pair and an "END_TITLE" pair to be respectively associated with the locations of the first and last words of the title."; col. 3, line 31, "means for indexing the parsed pages"; col. 5, lines 26-27, "In the index 70 each word is stored as a "literal" or character based value"; col. 8, lines 44-46, "By inserting the . "END_PAGE" attribute value in the index 70 as a metaword, searching the index as described below can be more efficient"; col. 9, lines 43-46, "the indexing module 40 generates an index 70 of the content of the records or pages 200. The internal data structures 71-73 of the index 70 are now described first with reference to FIG. 6");

said plurality of documents are de-duplicated in accordance with its fingerprint (col. 1, lines 42-45, Therefore, it is desired to provide a technique which minimizes the likelihood that duplicate pages are indexed. The technique should also allow for reindexing as duplicate pages are deleted."; col. 2, lines 41-42, "FIG. 24 shows a process for detecting duplicate pages; FIG. 25 is a flow diagram of a process for deleting pages"; col. 5, lines 12-14, "The maintenance module 80 also effectively deals with duplicate Web pages containing substantially identical content.");

said plurality of documents forming a cluster data base that is web-searchable by use of a predetermined descriptive term (col. 3, lines 28-33, "In order to identify pages of interest among the millions of pages which are available on the Web, a search engine 140 is provided. The search engine 140 includes means for parsing the pages, means

Art Unit: 2162

for indexing the parsed pages, means for searching the index, and means for presenting information about the pages 200 located.”).

However Burrows does not explicitly teach a system that stores electronic documents substantially equally throughout the cluster.

Narendran teaches a system that stores electronic documents substantially equally throughout the cluster (col. 17, lines 25-28, “the load distribution algorithm uses a load balancing metric to distribute the documents across the document servers such that request load is balanced across the document servers.”).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a system of storing electronic documents substantially equally throughout the cluster as disclosed by Narendran into the electronic document management system as disclosed in Burrows such that request load is balanced across the document servers (col. 17, lines 27-28). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

However the combination of Burrows and Narendran does not explicitly teach a system where each document is converted to ASCII text.

Froessler teaches a system where each document is converted to ASCII text (Abstract, “In one embodiment, the image representation is converted into code (ASCII)”).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate a system where each document is converted to

Art Unit: 2162

ASCII text as disclosed by Froessl into the an electronic document management system as disclosed in the combination of Burrows and Narendrant. Systems of this type allow full-text code searches to be conducted for words which appear in the documents. An advantage of this type of system is that indexing is not absolutely required because the full text of each document can be searched, allowing a document dealing with a specific topic or naming a specific person to be located without having to be concerned with whether the topic or person was named in the index (col. 1 lines 41-49). One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

Response to Arguments

9. Applicant's arguments regarding claims all pending claims have been fully considered but they are not persuasive.

Applicant's arguments regarding that Claim 1 is directed to "a method for managing a plurality of native documents to be uploaded to a document management computer system", and that Burrows is directed to the creation of an index of web pages, but not the management and distribution of the web pages, have been fully considered but they are not persuasive. In response to applicant's arguments, the recitation "a method for managing a plurality of native documents to be uploaded to a document management computer system" has not been given patentable weight because the recitation occurs in the preamble. A preamble is generally not accorded any patentable weight where it merely recites the purpose of a process or the intended use of a structure, and where the body of the claim does not depend on the preamble

for completeness but, instead, the process steps or structural limitations are able to stand alone. See *In re Hirao*, 535 F.2d 67, 190 USPQ 15 (CCPA 1976) and *Kropa v. Robie*, 187 F.2d 150, 152, 88 USPQ 478, 481 (CCPA 1951).

Applicant's arguments regarding that Burrows fails to disclose step (b) which recites "creating a fingerprint for each native document.", have been fully considered but they are not persuasive. Burrows explicitly discloses the creation of fingerprints for documents in column 8 lines 16-23, wherein "The fingerprint 255 can be produced by applying one-way polynomial functions to the digitized content".

Applicant's arguments regarding that Burrows fails to disclose de-duplicating each native document in accordance with the fingerprint, as claim 1 recites, have been fully considered but they are not persuasive. As admitted in the applicant's remarks, Burrows teaches that "... duplicate index entries are deleted" (page 15, argument 3, line 3), Burrows explicitly discloses the de-duplicating each native document from the index in accordance with the fingerprint as stated in the previous Office action, as well as the current Office action, in col. 1, lines 42-45, col. 2, lines 41-42, FIG. 24; FIG. 25, and col. 5, lines 12-14. When a document is de-duplicated from an index, a user is blocked from access to duplicate native documents, since a user's query will be run against the index; which contains only entries to unique native documents. For further support, see Burrows, col. 4, lines 49-50, "Users interact with the index 70 via the query module 50 by providing queries 52". Burrow's teaching reads on "de-duplicating each native document in accordance with the fingerprint" as claimed in Claim 1.

Applicant's arguments regarding that there is absolutely no disclosure or suggestion that Burrows is capable of distributing, or redistributing for that matter, millions of web pages that make up the internet amongst a plurality of nodes of a document management computer system, have been fully considered but they are not persuasive. As stated in this Office action, Burrows' teaching of a distributed system in FIG. 1 illustrates an environment of distributed servers, wherein in order for the clients to access information stored by the servers, the step of distributing documents on these servers is inherent, which may be further supported by the detailed description of Figure 1, col. 2 lines 52-56, "The distributed system 100 includes client computers 110 connected to server computers (sites) 120 via a network 130. The network 130 can use Internet communications protocols (IP) to allow the clients 110 to communicate with the servers 120", and col. 3 lines 1-9, "During operation of the distributed system 100, users of the clients 110 desire to access information records 122 stored by the servers 120 using, for example, the World-Wide-Web (WWW), or in short the "Web." The records of information 122 can be in the form of Web pages 200. The pages 200 can be data records including as content plain textual information, or more complex digitally encoded multimedia content, such as software programs, graphics, audio signals, videos, and so forth". Furthermore, in response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (for example., "distributing, or redistributing ..., millions of web pages that make up the internet...") are not recited in Claim 1. Although the claims are interpreted in light of the specification, limitations from the specification

are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

Applicant's arguments regarding that Getchius is absent any teaching of, especially distributing information substantially equally amongst a plurality of nodes, have been fully considered but they are not persuasive. The statement in the Patent Abstract, which the examiner relies on contains a typographical error "filly", which should have been corrected to read "fully". To further support this correction, see also col. 18 lines 41-43, "The architecture as depicted in FIGS. 2 and 4 includes a set of fully redundant server nodes in which each node is capable of responding to any search request". The fully redundant servers disclosed in Getchius reads on the limitation of distributing data substantially equally amongst a plurality of nodes. Getchius teaches redundant caching in col. 23, line 66 – col. 24 line 5, "Highly redundant caching is generally a technique that trades storage space against time by storing result sets along with subsets of these result sets. The highly redundant caching technique generally relies on the fact that the search time to locate an existing result is generally less than that amount of time which would result in creating the query result from a much larger search space". Based on the above teaching in Getchius that the set of servers are fully redundant and each node is capable of responding to any search request, it is inherent that same functionality and documents have been distributed to each of these server nodes, thus reading on the limitation of "distributing ... documents and ... data substantially equally amongst ... nodes...".

Furthermore, see also *In re Zletz*, 893 F.2d 319, 321-22, 13 USPQ2d 1320, 1322 (Fed. Cir. 1989) ("During patent examination the pending claims must be interpreted as broadly as their terms reasonably allow.... The reason is simply that during patent prosecution when claims can be amended, ambiguities should be recognized, scope and breadth of language explored, and clarification imposed.... An essential purpose of patent examination is to fashion claims that are precise, clear, correct, and unambiguous. Only in this way can uncertainties of claim scope be removed, as much as possible, during the administrative process.").

As reasons stated above for claim 1 in the previous as well as this office action, Examiner maintains that combination of Burrows and Getchius does teach a document management computer system as claimed in the applicant's invention.

Applicant's arguments regarding claims 6 and 7 that Burrows fails to disclose "discarding native documents that are determined to be the same in accordance with the comparison of fingerprints," have been fully considered but they are not persuasive. As stated in Applicant's Remarks, page 16, argument 7, line 3, "Burrows teaches discarding duplicate indexes." As reasons stated above for claim 1 in this office action, Examiner maintains that discarding duplicate indexes of documents read on discarding native documents that are determined to be the same. When a document is de-duplicated from an index, a user is blocked from access to duplicate native documents, since a user's query will be run against the index; which contains only entries to unique native documents. In other words, the duplicate native documents are discarded. See

also In re Zletz, 893 F.2d 319, 321-22, 13 USPQ2d 1320, 1322 (Fed. Cir. 1989)

("During patent examination the pending claims must be interpreted as broadly as their terms reasonably allow.... The reason is simply that during patent prosecution when claims can be amended, ambiguities should be recognized, scope and breadth of language explored, and clarification imposed.... An essential purpose of patent examination is to fashion claims that are precise, clear, correct, and unambiguous. Only in this way can uncertainties of claim scope be removed, as much as possible, during the administrative process.").

Applicant's arguments regarding claims 8 and 9 that Burrows fails to disclose or suggest creating at least one datafile or a plurality of datafiles corresponding to the extracted data for each native document, have been fully considered but they are not persuasive. Burrows teaches the summary data structure in column 11, line 66-column 12, line 7, the generation of summary entries in the second level summary data structure whereby each summary entry includes words associated with the sample. Samples are analogous to extracted data and summaries are analogous to data files generated based on the samples. For further support for the sampling concept, see Burrows col. 11, lines 51-60. Therefore, the examiner maintains that Burrows discloses or suggests creating at least one datafile or a plurality of datafiles corresponding to the extracted data for each native document as claimed.

Applicant's arguments regarding claim 29 that the network of Burrows is not maintained in accordance with the invention described by claim 1. Rather, Burrows is merely directed to the creation of an index of the web pages on the computer system, have been fully considered but they are not persuasive. In response to applicant's argument that Burrows is merely directed to the creation of an index of the web pages on the computer system, a recitation of the intended use of the claimed invention must result in a structural difference between the claimed invention and the prior art in order to patentably distinguish the claimed invention from the prior art. If the prior art structure is capable of performing the intended use, then it meets the claim. In a claim drawn to a process of making, the intended use must result in a manipulative difference as compared to the prior art. See *In re Casey*, 370 F.2d 576, 152 USPQ 235 (CCPA 1967) and *In re Otto*, 312 F.2d 937, 939, 136 USPQ 458, 459 (CCPA 1963). . As stated in the Office action that Burrows (col. 2, lines 51-56) discloses a system comprising a computer in communication with the plurality of computer nodes as claimed in the applicant's invention.

Applicant's arguments regarding claim 2 that the motivation provided by the Examiner is mere hindsight and does not, in any way, establish any motivation for modifying Burrows, which directed to a indexable web server and that Burrows does not disclose the ability to index archived files and hence, there is no support or rational link for the combination suggested, have been fully considered but they are not persuasive. In response to applicant's argument that the examiner's conclusion of obviousness is

based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

In response to applicant's argument that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988) and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, Burrows' document indexing system can be easily modified to include documents extracted from an archive database as disclosed in Okabe. Documents extracted from an archive database is no different from a normal native document. Both One of ordinary skill in the art would be motivated to make the aforementioned combination with reasonable expectation of success.

Applicant's arguments regarding claim 3 that the Examiner is using hindsight reasoning to arrive at the proposed combination, have been fully considered but they are not persuasive. In response to applicant's argument that the examiner's conclusion

of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

In response to applicant's argument that Burrows (or Getchius) fails to teach creating a fingerprint, only detecting a fingerprint and hence, there is no support for the combination suggest, the examiner relies on the response stated above for claim 1, regarding Burrows' teaching on creating a fingerprint for each native document. Both Burrows reference and Zabetian reference teach creating a fingerprint for a document, and thus both belong to the same field of endeavor. It would be obvious for one of ordinary skill in the art to borrow features advantageous from one another.

Applicant's arguments with respect to Claim 34 have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

10. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Bestavros, "Demand-based document dissemination to reduce traffic and balanceload in distributed information systems": a technique is provided to considerably

Art Unit: 2162

reduce network traffic and minimize the latency of information retrieval in a distributed system by disseminating the most popular documents on servers closer to clients, while servers are load-balanced.

11. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the date of this final action.

Contact Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to GWEN LIANG whose telephone number is 571-272-4038. The examiner can normally be reached on 9:00 A.M. - 5:30 P.M. Monday and Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JOHN BREENE can be reached on 571-272-4107. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

G.L.
7 June 2005

John S. Kossman
Primary Examiner